Imagining sounds and images: Decoding the contribution of unimodal and transmodal brain regions to semantic retrieval in the absence of meaningful input

Abbreviated Title – Imagining sounds and images

Charlotte Murphy[1], Shirley-Ann Rueschemeyer[1], Jonathan Smallwood[1] and Elizabeth Jefferies[1]

[1] Department of Psychology / York Neuroimaging Centre, University of York.

Address for correspondence:
Charlotte Murphy
Department of Psychology / York Neuroimaging Centre, University of York. Email: charlotte.murphy@york.ac.uk
Tel +44 (0)1904 323190

Abstract

In the absence of sensory information, we can generate meaningful images and sounds from representations in memory. However, it remains unclear which neural systems underpin this process, and whether different types of imagery recruit similar or different neural networks. We asked people to imagine the visual and auditory features of objects, either in isolation (car, dog) or in specific meaning-based contexts (car/dog race). Using an fMRI decoding approach, in conjunction with functional connectivity analysis, we examined the role of primary auditory/visual cortex and transmodal brain regions. Conceptual retrieval in the absence of external input recruited sensory and transmodal cortex. The response in transmodal regions – including anterior middle temporal gyrus – was of equal magnitude for visual and auditory features, yet nevertheless captured modality information in the pattern of response across voxels. In contrast, sensory regions showed greater activation for modality-relevant features in imagination (even when external inputs did not differ). These data are consistent with the view that transmodal regions support internally-generated experiences and that they play a role in integrating perceptual features encoded in memory.

Introduction

In the absence of sensory information, the mind produces experiences with rich sensorimotor features through the retrieval of information from memory (Singer, 1966; Antrobus, Singer & Greenberg, 1966; Mason et al. 2007). For instance, in everyday life we regularly hear voices and music in the mind's ear when no sound is delivered (e.g., Alderson & Fernyhough, 2015; Halpern, 2001) and studies suggest more than one third of our time is spent engaged in thoughts and experiences that are unrelated to the ongoing environment (Kane et al. 2007; Killingsworth & Gilbert, 2010). Although attempts have been made to understand how the brain retrieves memories in the absence of input (Albers et al. 2013; Daselaar, Porat, Huijbers & Pennartz, 2010; Vetter, Smith & Muckli, 2014), we lack an account of the component neurocognitive processes critical for mental imagery, whether these vary with respect to the modality of the memories being retrieved, and how these processes combine to support more complex multi-dimensional aspects of cognition. Studies of imagination have almost entirely focused on a constrained regions-of-interest analysis, which may not adequately represent the rich involvement of multiple brain regions distributed across the cortex. Moreover, they have seldom attempted to differentiate between different forms of imagery, with the majority of studies focusing solely on visual imagery (Albers et al. 2013; Countanche & Thompson-Schill, 2014; Dijkstra et al. 2017; Ishai, Ungerleider & Haxby, 2000; Lee, Kravitz & Baker, 2012; Reddy, Tsuchiya & Serre, 2010; Stokes, Thompson, Cusack & Duncan, 2009; Vetter et al. 2014). As such, there is limited understanding of the neural signature of different modalities (e.g., visual versus auditory), and whether different forms of imagination share similar or unique neural representations. Notably, studies that have compared visual and auditory imagery within the same experiment have been criticized for not employing comparable task conditions (see Daselaar et al. 2010; Halpern et al. 2004).

We addressed these issues by applying multivoxel pattern analysis (MVPA) and resting-state functional magnetic resonance imaging to identify neural patterns that support different aspects of imagination at the whole-brain level. Using a

constant source of visual and auditory noise as a baseline, participants were asked to imagine information under three different conditions: visual (e.g. what a dog looks like), auditory (what a dog sounds like) and contextual (e.g. imagining a dog in a specific context, such as a race dog). This latter condition combines features from multiple modalities in a complex way (e.g., imagining a race dog may involve the visual properties of a greyhound and race track, as well as the auditory properties of dogs panting and crowds cheering). Figure 1 presents a schematic description of the experimental design used in our experiment. We compared the time points during which participants imagined a given concept whilst observing visual and auditory noise to those in which participants only observed visual and auditory noise (baseline). Our paradigm, therefore, permitted us to investigate the mechanisms involved in imagery whilst controlling for sensory input across our conditions.

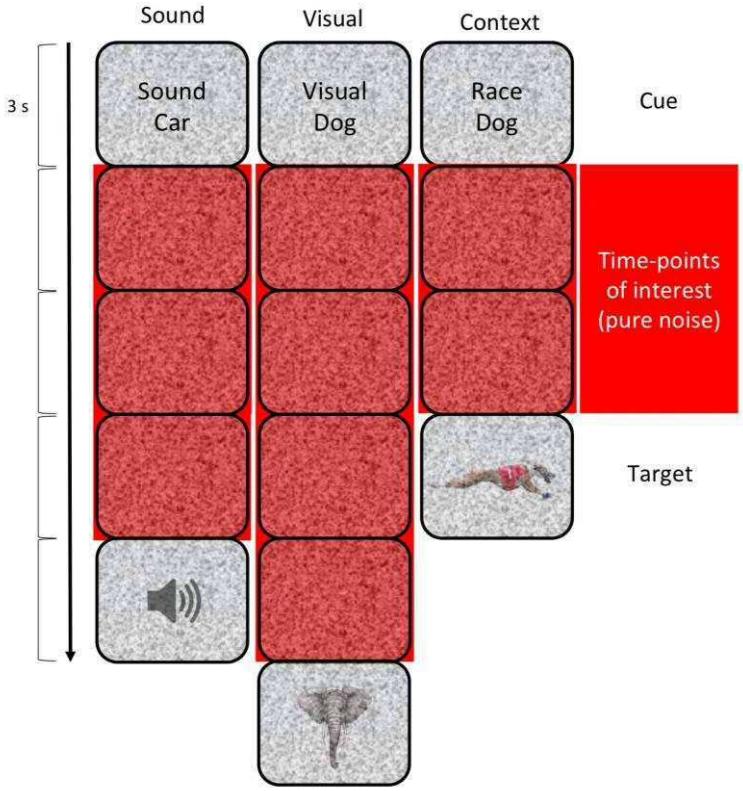

Figure 1. Experimental design. Participants were presented with written cues embedded in visual and auditory noise that referred to items they must detect. Cues referred to one of three tasks (Thinking about the *sound* of a concept; Thinking about the *visual* properties of a concept; Thinking about a concept in a particular complex *context* i.e., at the races) for one of two concepts (Dogs; Cars). This yielded six experimental conditions (Sound Car; Sound Dog; Visual Car; Visual Dog; Context Car (e.g., Race Car); Context Dog (e.g., Race Dog)). Cues

were followed by blocks of pure noise that lasted 6-12 s. Each block ended with either an image or a sound embedded in noise, that was either congruent to the cue (e.g., greyhound for the context cue 'Race Dog') or incongruent (e.g., elephant trunk for the visual cue 'Visual Dog'). Participants responded with a yes/no response to whether the target trial matched the cue. Time points of interest are highlighted in red, these refer to pure noise trials where participants were thinking about the relevant cue (e.g., thinking about what a sound looked like). Cues, each pure-noise image and targets were shown for 3 s each.

A wealth of evidence supports the view that regions of unimodal sensory cortex are important for modality-specific elements of memory retrieval during imagination. Visual cortex is activated by mental images (Albers et al. 2013; de Borst & de Gelder, 2016; Ishai et al. 2000; Reddy et al. 2010; Vetter et al. 2014) and auditory cortex is activated by imagined sounds (Daselaar et al. 2010; de Borst & de Gelder, 2016; Halpern & Zatorre, 1999; Zvyagintsev et al. 2013). These findings are consistent with embodied cognition accounts, which propose that sensory regions important for perception and action also support mental processes such as comprehension and imagery (for discussion, see Barsalou, 1999; 2008; Patterson, Nestor & Rogers, 2007; Kiefer & Pulvermüller, 2012). Notably, the majority of studies find recruitment of sensory association cortices during visual (Amedi et al. 2005; Ishai et al. 2000; Knauff et al. 2000) and auditory mental imagery (Bunzeck et al. 2005; Zatorre & Halpern, 2005). Moreover, a recent fMRI study showed that both secondary sensory regions and top-down mechanisms are necessary in visual imagery for enhancing the relevant representations in early sensory areas (Dijkstra et al. 2017). Some studies have also found imagery-induced activation in primary sensory cortex (Kosslyn et al. 1999; 2001; Slotnick, Thompson & Kosslyn, 2005), and the extent to which primary and/or secondary sensory regions are recruited during different modalities of imagery remains a source of contention (Daselaar et al. 2010; Kosslyn et al. 2001). By directly comparing visual and auditory imagery under equivalent conditions in the same experiment, the present study can elucidate the role of sensory cortex in mental imagery.

Contemporary accounts of semantic cognition suggest that memory retrieval also relies on abstract representations that are largely invariant to the input modality. A prominent theory of conceptual representation, known as the hub-and-spoke account, suggests that input-invariant concepts draw on a convergence zone

in the ventrolateral anterior temporal lobes (ATL), which extracts deep semantic similarities across multiple unimodal features (Lambon Ralph, Jefferies, Patterson & Rogers, 2017; Patterson et al. 2007). Support for this account comes from a recent fMRI study utilizing MVPA, which demonstrated that anterior inferior and middle temporal gyrus support modality-invariant patterns of activity corresponding to meaning. In contrast, superior temporal voxels held patterns of activity that reflected sensory input modality (Murphy et al. 2017). If ventrolateral ATL represents abstract conceptual representations, as expected for a transmodal brain region (Margulies et al. 2016; Mesulam, 2012), it may be critical for stimulus-independent cognition regardless of the modality that is being imagined.

In line with this broad perspective, studies have revealed ventrolateral ATL activation during the retrieval of concepts across different input modalities (e.g., Gabrieli et al. 1997; Murphy et al. 2017; Reilly, Garcia & Binney, 2016; Rice et al. 2015; Van Ackeren & Rueschemeyer, 2014; Visser et a. 2011). Coutanche and Thompson-Schill (2014) also found that left ATL could successfully decode the properties of an imagined object. In this study, classifiers in visual regions related to the shape (in V1) and colour (in V4) of the object predicted classification of the specific imagined object in ATL. This is consistent with the hypothesis that information from sensory cortex is integrated in ATL to form modality-invariant conceptual representations that are critical for perceptually-decoupled semantic retrieval. However, this previous study only examined visual features, while connectivity and task activation data suggest ATL acts as a convergence zone across different sensory modalities, including auditory features (Patterson et al., 2007; Visser et al., 2010; Lambon Ralph et al., 2017). Since the convergence of these different modalities is thought to be graded (Lambon Ralph et al, 2017), it is assumed that ventrolateral ATL retains some degree of differential connectivity to auditory and visual cortex. A key question, therefore, is whether transmodal portions of ATL play a common or distinct role in the representation of information about different modalities in imagination (e.g., when imagining visual and auditory features in the absence of input).

Furthermore, our context condition (race + dog) permits us to investigate brain regions recruited during more complex multimodal imagery (e.g., imagining a

dog race may involve the visual properties of a greyhound and race track, as well as the auditory properties of the dogs thundering down the track and crowds cheering). Baron and Osheron (2011) found that conceptual combinations were represented in left ATL: decoding accuracy was related to classification accuracy for the constituents (boy = young + man). ATL can also show a stronger response to conceptual combinations, perhaps because these combinations require more specific patterns of semantic retrieval (Bemis & Pylkkanen, 2012). However, recent studies have shown that complex mental events are associated with a broader transmodal network including medial prefrontal cortex (Hartung et al. 2015) and attentional mechanisms (Berger, 2016). Taken together this literature suggests that the heteromodal regions recruited to support simple semantic imagery across visual and auditory features may not be sufficient when imagination is more complex: additional regions may come into play to support our capacity to flexibly maintain and integrate multiple features in specific and diverse ways.

The present study used a combination of imaging methods to understand patterns of common and distinct neural activity that are important for different forms of mental imagery (auditory features, visual features and complex conceptual combinations). First, we used MVPA to identify regions that code for each condition. Second, we performed conjunctions of these MVPA maps to identify distinct regions representing the presence or absence of a specific condition. Third, we interrogated the univariate activation of our conjunction maps to identify the BOLD response in each region. Fourth, we seeded these maps in an independent resting-state cohort to identify the intrinsic networks that these fall within. Finally, we performed a conjunction of these resting-state maps to identify potential common regions within the large-scale networks necessary for all forms of imagery. To complement these resting state analyses, we performed a meta-analysis of these spatial maps to provide a quantitative description of the types of cognitive processes that these patterns are linked to.

Using this analysis pipeline, the present study examined three questions that emerge from a common and distinct account of semantic retrieval in the absence of meaningful input. First, we examined whether different types of sensory cortex play a specific role in memory retrieval. For example, auditory cortex should be recruited

more for thinking about what a dog sounds like than what it looks like; moreover the patterns of activity in this region should be able to decode between thinking about auditory features and other forms of imagery (e.g., visual or context conditions). Given that the majority of the literature highlights the recruitment of sensory association cortex, we predicted that secondary sensory regions would be recruited more extensively than primary sensory regions during imagery. Second, we investigated the contribution of transmodal regions, including ATL, to different forms of imagery. If these regions combine information from different modalities in a *graded* fashion, differential connectivity might allow these regions to classify imagined visual and auditory features. Finally, using resting-state fMRI, we characterized the intrinsic connectivity of regions identified in our MVPA analysis to understand the neural networks they are embedded in. We anticipated that these regions would show functional connectivity to regions of transmodal cortex implicated in abstract forms of cognition, as well as to relevant portions of sensory cortex (i.e. visual cortex during visual imagery). Together these different analytic approaches permit the investigation of both similarities and differences in the networks recruited when semantic retrieval is internally-generated.

Materials and Methods

Functional Experiment

Participants. Twenty participants were recruited from the University of York. One participant's data was excluded due to excessive motion artifacts, leaving nineteen subjects in the final analysis (11 female; mean age 23.67, range 18-37 years). Participants were native British speakers, right handed and had normal or corrected-to-normal vision. Participants gave written informed consent to take part and were reimbursed for their time. The study was approved by the York Neuroimaging Centre Ethics Committee at the University of York.

Design. The functional experiment contained six experimental conditions, in a 2 (concepts; dog, car) x 3 (type of imagery; auditory, visual and conceptually-complex context) design (see supplementary table A2 for full list of experimental conditions).

Stimuli. Experimental stimuli consisted of (i) six verbal conceptual prompts that referred to each of our six experimental conditions (e.g., Dog Sound, which cued participants to imagine what a dog sounded like), (ii) visual and auditory noise which was presented throughout experimental conditions and rest periods. For this, Gaussian visual noise was generated through Psychopy (Pyschopy, 2.7), and auditory white noise was generated through Audacity software (Audacity Version 2.0.0), and (iii) target images/sounds. The targets used in this paradigm were piloted prior to fMRI scanning, on a separate group of participants (n=24) to determine the average length of time taken to detect a target (image or sound) emerging through noise (see supplementary material A1 for full description of pilot experiment). From this pilot, ten images were selected for each of our six experimental conditions (Dog Visual-Features, Car Visual-Features, Dog Sound, Car Sound, Dog Context and Car Context) based on statistically similar reaction times (RTs) for detecting the item emerging through noise. Images were detected on average at 2861 ms and sounds at 2912 ms (see Table 1). The fMRI scan therefore allowed 3000 ms for participants to detect whether an item emerging through noise matched the content of their imagery.c;t

Task Procedure. Prior to being scanned participants completed a practice session, identical to one scanning run. After this practice run, participants were probed to describe what they had been imagining during the pure noise trials to ensure the participants were engaged in imagining the relevant concepts. For the in-scanner task stimuli were presented in four independent runs. Within each scanning run participants were presented a cue word (e.g., Sound DOG) and instructed to imagine this concept in the presence of visual and auditory noise; for instance, they could imagine the sound of a dog barking, growling, yelping etc. Task instructions were presented for 3s. A variable number of images then followed, each displaying visual and auditory noise (see Figure 1). Within the blocks, the pure-noise images were each shown for 3s. Following a variable length of time (between 6 and 12s after the

initial cue), a target image or sound began to emerge through the noise (at the rate outlined in the pilot experiment described above). Participants were instructed to respond with a button-press (yes/no) whether a target item emerging through visual and auditory noise was related to what they had being imagining based on the cue word. Participants were given 3000ms to respond to this item. The block automatically ended after this image. This design afforded us the high signal sensitivity found with block designs, combined with unpredictability to keep participants cognitively engaged.

Each experimental condition (e.g., "Dog Sound") occurred two times in a run (giving 8 blocks for each experimental condition across the experiment). Blocks were presented in a pseudo-randomized order so the same cue did not immediately repeat, and blocks were separated by 12s fixation. During the fixation period the visual and auditory noise were also presented, to create an active baseline. 50% of the items emerging through noise contained an item that did not match the preceding cue (i.e., 4 of 8 were foils) in order to focus participants on detecting the specific target. To encourage participants to pay attention from the very start of every block, an additional short block was included in each run, in which an item emerged through noise after only 3s, followed by 12s of fixation. These blocks were disregarded in the analysis.

Acquisition. Data were acquired using a GE 3T HD Excite MRI scanner at the York Neuroimaging Centre, University of York. A Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage, radio-frequency coil tuned to 127.4MHz. A gradient-echo EPI sequence was used to collect data from 38 bottom-up axial slices aligned with the temporal lobe (TR = 2s, TE = 18 ms, FOV = 192 × 192 mm, matrix size = 64 × 64, slice thickness = 3 mm, slice-gap 1mm, flip-angle = 90°). Voxel size was 3 × 3 × 3 mm. Functional images were co-registered onto a T1-weighted anatomical image from each participant (TR = 7.8 s, TE = 3 ms, FOV = 290 mm x 290 mm, matrix size = 256 mm x 256 mm, voxel size = 1.13 mm x 1.13 mm x 1 mm) using linear registration (FLIRT, FSL). This sequence was chosen as previous studies employing this sequence have produced an adequate signal-to-noise ratio in

regions prone to signal dropout, such as ATL (e.g., Coutanche & Thompson-Schill, 2014; Murphy et al. 2017).

To ensure that our ROIs had sufficient signal to detect reliable fMRI activation, the temporal signal-to-noise ratio (tSNR) for each participant was calculated by dividing the mean signal in each voxel by the standard deviation of the residual error time series in that voxel (Friedman et al., 2006). tSNR values were averaged across the voxels in both anterior temporal lobe (ATL) and medial prefrontal cortex (mPFC); regions that suffer from signal loss and distortion due to their proximity to air-filled sinuses (Jezzard & Clare, 1999). Mean tSNR values, averaged across participants, were as follows: ATL, 82.85; mPFC, 97.14. The percentage of voxels in each ROI that had "good" tSNR values (>20; Binder et al., 2011) was above 97% for all ROIs: ATL, 97.19%; mPFC, 99.24%. These values indicate that the tSNR was sufficient to detect reliable fMRI activation in all ROIs (Binder et al., 2011).

Pre-processing. Imaging data were preprocessed using the FSL toolbox (http://www.fmrib.ox.ac.uk/fsl). Images were skull-stripped using a brain extraction tool (BET, Smith, 2002) to remove non-brain tissue from the image. The first five volumes (10s) of each scan were removed to minimize the effects of magnetic saturation, and slice-timing correction was applied. Motion correction (MCFLIRT, Jenkinson et al. 2002) was followed by temporal high-pass filtering (cutoff = 0.01 Hz). Individual participant data were first registered to their high-resolution T1-anatomical image, and then into a standard space (Montreal Neurological Institute (MNI152); this process included tri-linear interpolation of voxel sizes to 2 × 2 × 2 mm. For univariate analyses, data were additionally smoothed (Gaussian full width half maximum 6 mm).

Multivariate Pattern Analysis. Analysis was focused on the moments when participants were imagining the target cues (e.g., thinking about what a dog looked like, or what a car sounded like). The condition onset and duration were taken from the first pure noise trial in each block (after the initial cue) to the end of the last pure noise trial (before the item began to emerge through the noise). The response to

each of the 6 conditions was contrasted against the active rest baseline (periods of auditory and visual noise where participants were not cued to imagine concepts). Box-car regressors for each condition, for each run, in the general linear model were convolved with a double gamma hemodynamic response function (FEAT, FSL). Regressors of no interest were also included to account for head motion within scans. MVPA was conducted on spatially unsmoothed data to preserve local voxel information. For each voxel in the brain, we computed a linear support vector machine (LIBSVM; with fixed regularization hyper-parameter C = 1) and a 4-fold cross-validation (leave-one-run-out) classification, implemented in custom python scripts using the pyMVPA software package (Hanke et al. 2009). A support vector machine was chosen to combat over-fitting by limiting the complexity of the classifier (Lewis-Peacock & Norman, 2013). The classifier was trained on three runs and tested on the independent fourth run; the testing set was then alternated for each of four iterations. Classifiers were trained and tested on individual subject data transformed into MNI standard space. The functional data were first z-scored per voxel within each run. The searchlight analysis was implemented by extracting the z-scored β-values from spheres (6mm radius) centered on each voxel in the masks. This sized sphere included

with previous decoding studies of internally generated thought, which have shown that specific-level concepts (e.g., lime vs. celery) can be decoded; however categorical-level concepts (e.g., fruit vs. vegetable) were not successfully classified (Coutanche & Thompson-Schill, 2014). This may reflect the dynamic nature of conceptually driven internally-generated thought; for instance, on one trial, subjects may have been thinking about the exterior look of a car and on the next trial imagining the interior decor. As this analysis revealed no regions across the cortex could successfully decode this information, the remaining classification tests combine car and dog trials. (2) Auditory vs. visual classifier: this examined whether patterns of activity conveyed information regarding the modality of imagery, by training a classifier to discriminate between periods of noise where participants were thinking about the visual properties of objects and periods of noise where participants were thinking about the auditory properties of objects. (3) Visual vs. context classifier: here a classifier was trained to discriminate between periods of noise where participants were thinking about the visual properties of objects and periods of time when participants were thinking about objects in more complex conceptual contexts. (4) Auditory vs. context classifier: here a classifier was trained to discriminate between periods of noise where participants were thinking about the auditory properties of objects and period of time when participants were thinking about objects in complex contexts. Unthresholded maps from all analyses are uploaded on Neurovault: http://neurovault.org/collections/2671/.

Next, we identified regions where patterns of activity consistently informed the classifier for each of our three tasks (visual, auditory and context) by running a formal conjunction on the uncorrected searchlight maps (using the FSL easythresh command). For visual patterns we looked at the conjunction of the two searchlight maps that decoded visual properties (visual vs. auditory and visual vs. context). Since regions that contributed to both of these searchlight maps were able to decode simple visual features in imagination, relative to both auditory features and more complex contexts, we reasoned that their pattern of activation related to simple visual features. Next, we looked at the conjunction of the two searchlight maps that decoded the auditory condition (auditory vs. visual and auditory vs. context), to identify brain regions containing patterns of activation relating to simple auditory

properties in imagination. Finally, we looked at the conjunction of the two searchlight maps that decoded context properties (context vs. visual and context vs. auditory). This identified brain regions containing activation patterns relating to complex conceptual contexts, as distinct from both simple visual and auditory features. All analyses were cluster corrected using a z-statistic threshold of 2.3 to define contiguous clusters. Multiple comparisons were controlled using a Gaussian Random Field Theory at a threshold of $p < .01$.

Univariate Analysis. We examined univariate activation to further characterise the response within our unimodal and transmodal regions defined by MVPA. The percent signal change was extracted for each condition from regions of interest (ROIs) defined by the MVPA conjunctions (see above).

Resting state fMRI

Participants. This analysis was performed on a separate cohort of 157 healthy participants at York Neuroimaging Centre (89 female; mean age 20.31, range 18–31 years). Subjects completed a 9-minute functional connectivity MRI scan during which they were asked to rest in the scanner with their eyes open. Using these data, we examined the resting-state fMRI (rs-fMRI) connectivity of our conjunction regions that were informative to decoding visual imagery, auditory imagery and contextual imagery, to investigate whether these regions fell within similar or distinct networks. The data from this resting-state scan has been used in prior published works from the same lab (e.g., Murphy et al. 2017; Murphy et al. 2018; Vatansever et al. 2017; Wang et al. 2018).

Acquisition. As with the functional experiment, a Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage, radio-frequency coil tuned to 127.4 MHz. For the resting-state data, a gradient-echo EPI sequence was used to collect data from 60 axial slices with an interleaved (bottom-up) acquisition order with the following parameters: TR=3 s, TE=minimum full, volumes=180, flip angle=90°, matrix size=64×64, FOV=192×192 mm, voxel size=3x3×3 mm. A minimum

full TE was selected to optimise image quality (as opposed to selecting a value less than minimum full which, for instance, would be beneficial for obtaining more slices per TR). Functional images were co-registered onto a T1-weighted anatomical image from each participant (TR=7.8 s, TE=3 ms, FOV=290 mmx290 mm, matrix size=256 mm x256 mm, voxel size=1 mm x 1 mm x 1 mm).

Pre-processing. Data were pre-processed using the FSL toolbox (http://www.fmrib.ox.ac.uk/fsl). Prior to conducting the functional connectivity analysis, the following pre-statistics processing was applied to the resting state data; motion correction using MCFLIRT to safeguard against motion-related spurious correlations slice-timing correction using Fourier-space time-series phase-shifting; non-brain removal using BET; spatial smoothing using a Gaussian kernel of FWHM 6 mm; grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative factor; high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma=100s); Gaussian low-pass temporal filtering, with sigma=2.8s.

Low-level analysis. For each conjunction site we created spherical seed ROIs, 6mm in diameter, centered on the peak conjunction voxel; visual conjunction site in left inferior lateral occipital cortex (LOC) [-48 -70 -2], auditory conjunction site in left superior temporal gyrus [-48 -12 -10] and context conjunction site in left LOC [-48 -60 0] respectively (see supplementary table A2). This ensured that we assessed the functional connectivity of a key site when the searchlight conjunction revealed a large cluster or multiple clusters. The time series of these regions were extracted and used as explanatory variables in a separate subject level functional connectivity analysis for each seed. Subject specific nuisance regressors were determined using a component based noise correction (CompCor) approach (Behzadi et al. 2007). This method applies principal component analysis (PCA) to the fMRI signal from subject specific white matter and CSF ROIs. In total there were 11 nuisance regressors, five regressors from the CompCor and a further 6 nuisance regressors were identified using the motion correction MCFLIRT. These principle components were then removed from the fMRI data through linear regression. The WM and CSF covariates

were generated by segmenting each individual's high-resolution structural image (using FAST in FSL; Zhang et al. 2001). The default tissue probability maps, referred to as Prior Probability Maps (PPM), were registered to each individual's high-resolution structural image (T1 space) and the overlap between these PPM and the corresponding CSF and WM maps was identified. These maps were then thresholded (40% for the SCF and 66% for the WM), binarized and combined. The six motion parameters were calculated in the motion-correction step during pre-processing. Movement in each of the three Cartesian directions (x, y, z) and rotational movement around three axes (pitch, yaw, roll) were included for each individual.

High-level analysis. At the group-level the data were processed using FEAT version 5.98 within FSL (www.fmrib.ox.ac.uk/fsl) and the analyses were carried out using FMRIB's Local Analysis of Mixed Effects (FLAME) stage 1 with automatic outlier detection. No global signal regression was performed. The z statistic images were then thresholded using clusters determined by $z > 2.3$ and a cluster-corrected significance threshold of $p = 0.05$. Finally, to determine whether our connectivity maps overlapped with one another we calculated the number of overlapping voxels for our three conjunction site connectivity maps.

Results

Behavioural Results

To determine whether our experimental conditions were well matched at the behavioural level, accuracy and reaction times (RT) for the fMRI session were calculated for each participant (n=19). All participants were engaged in the correct task (e.g., thinking about the sound of a dog) as indicated by a mean accuracy score above 75% for all experimental conditions (Table 1). A 2 (semantic category; car, dog) by 3 (imagery type; auditory, visual, context) repeated-measures ANOVA revealed no differences in accuracy across the three types of imagery (auditory, visual, conceptually-complex context; $F(2,36) = 2.32$, $p = .11$) and no effect of concept (car, dog; $F(1,18) = 1.95$, $p = .66$). RT scores were also well matched across our experimental conditions (Table 1). A 2 x 3 repeated measures ANOVA revealed

there was no difference in RT between the three experimental tasks (auditory, visual, conceptually-complex context; $F(2,36) = 0.46$, $p=.64$), no effect of concept (car, dog; $F(1,18) = 2.61$, $p=.09$) and no interaction between imagery types and concept ($F(2,36) = 1.17$, $p = .37$). Furthermore, the in-scan RT data were close to the RT in our pilot study, suggesting that participants required the same amount of time to detect stimuli both in and out of the scanner (mean RT for images = 2660 ms, SD = 233 ms, mean RT for sounds = 2763 ms, SD = 616 ms).

Table 1. Behavioural scores across pilot and fMRI experiments

Condition	Pilot Experiment		fMRI Experiment	
	RT	Acc	RT	Acc
Car Sound	2873 (635)	N/A	2748 (713)	82.11 (16.53)
Dog Sound	2951 (876)	N/A	2753 (552)	76.84 (12.04)
Car Visual	2886 (367)	N/A	2704 (204)	83.68 (11.64)
Dog Visual	2812 (402)	N/A	2620 (241)	82.63 (9.91)
Car Context	2994 (355)	N/A	2754 (211)	76.76 (12.62)
Dog Context	2752 (398)	N/A	2569 (250)	79.61 (14.71)

Footnote: RT = reaction time in milliseconds, ACC = percentage accuracy. Standard deviation in parentheses.

MVPA Decoding Results

To test which brain regions held patterns of activity related to the type of internally-generated conceptual retrieval, we examined brain regions that could classify imagery conditions during the presentation of auditory and visual noise. For example, the auditory vs. visual classifier was trained on the distinction between thinking about auditory and visual properties of concepts (collapsed across both cars and dogs) and tested on the same distinction in unseen data using a cross-validated approach. All results reported are above chance levels (50%, cluster-corrected, $p < .01$).

The whole-brain searchlight analysis for the distinction between visual and auditory imagery revealed an extensive network of brain regions including sensory regions, such as bilateral inferior lateral occipital cortex (LOC), left fusiform and left auditory cortex (encompassing planum polare and Heschl's gyrus extending more broadly into superior temporal gyrus), as well as transmodal brain regions that have been implicated in semantic processing, such as middle temporal gyrus, ATL (middle, inferior, fusiform and parahippocampal portions) and on the medial surface, anterior cingulate gyrus and thalamus (see Figure 2A; Table 2).

Table 2. Centre Voxel Coordinates of Highest Decoding Sphere in the Searchlight Analyses.

Condition	Cluster Peak	Extended Cluster Regions	Cluster Extent	Z-Score	Acc (%)	x	y	z
Auditory vs. Visual								
	L Lateral occipital cortex, superior division	L Lateral occipital cortex, inferior division, L Occipital pole, L Occipital fusiform gyrus.	975	4.13	75.00%	-36	-86	10
	L Thalamus	R Thalamus	599	4.18	66.25%	-12	-26	2
	R Lateral occipital cortex, inferior division	R Middle temporal gyrus, temporooccipital part.	431	4.43	68.75%	54	-66	10
	L Planum polare	L Superior temporal gyrus, posterior division , Insular cortex, L Heschl's gyrus, Anterior superior temporal gyrus.	226	3.77	70.75%	-40	-16	-8
	L Supramarginal gyrus, posterior division	L Planum temporale, Posterior superior temporal gyrus.	178	3.52	75.00%	-60	-42	16
	R Frontal operculum cortex	R Frontal orbital cortex, R Insular cortex.	156	3.37	68.25%	40	22	4
	L Anterior parahippocampal gyrus	L Temporal fusiform gyrus, ,	75	4.34	75.00%	-36	-18	-18
	L Anterior middle temporal gyrus	L Anterior inferior temporal gyrus,	67	4.12	66.25%	-56	-6	-18
	L Anterior cingulate gyrus		49	3.82	58.36%	-4	34	-2

Visual vs. Context								
	L Lateral occipital cortex, inferior division	L Middle temporal gyrus, temporooccipital part, L Occipital Pole.	733	4.16	68.75%	-46	-72	0
Auditory vs. Context								
	L Lateral occipital cortex, inferior division	L Temporal occipital fusiform cortex, L inferior temporal gyrus, temporooccipital part.	312	3.81	76.49%	48	-62	-6
	R Temporal occipital fusiform gyrus	R Lateral occipital cortex, inferior division, R Inferior temporal gyrus, temporooccipital part, R Middle temporal gyrus. temporooccipital part	118	3.17	68.75%	34	-56	-20
	R Posterior middle temporal gyrus	R Posterior superior temporal gyrus, R Supramarginal gyrus, R Anterior superior temporal gyrus	90	2.92	68.75%	56	-34	-2
	R Posterior superior temporal gyrus	R Middle temporal gyrus, R Planum polare, R Planum Temporale	81	3.15	75.00%	60	-22	0

Footnote: Highest decoding accuracy clusters for each of our three classifiers analysed separately. The Auditory vs. Visual classifier was trained on the distinction between thinking about the sound of a concept versus thinking about what a concept looked like. The Visual vs. Context classifier was trained on the distinction between thinking about what a concept looked like versus thinking about it in a specific meaning-based context. The Sound vs. Context classifier was trained on the distinction between thinking about what a concept sounded like and thinking about it in a specific meaning-

based context. All analyses were cluster corrected using a z-statistic threshold of 2.3 to define contiguous clusters. Multiple comparisons were controlled using a Gaussian Random Field Theory at a threshold of $p < .01$. L = left, R = right. As well as peak accuracy (reported under the 'Cluster Peak' column), the 'Extended Cluster Regions' includes all significant regions within each ROI. The unthresholded MVPA maps for each searchlight have been uploaded to the Neurovault database and can be found here http://neurovault.org/collections/2671/.

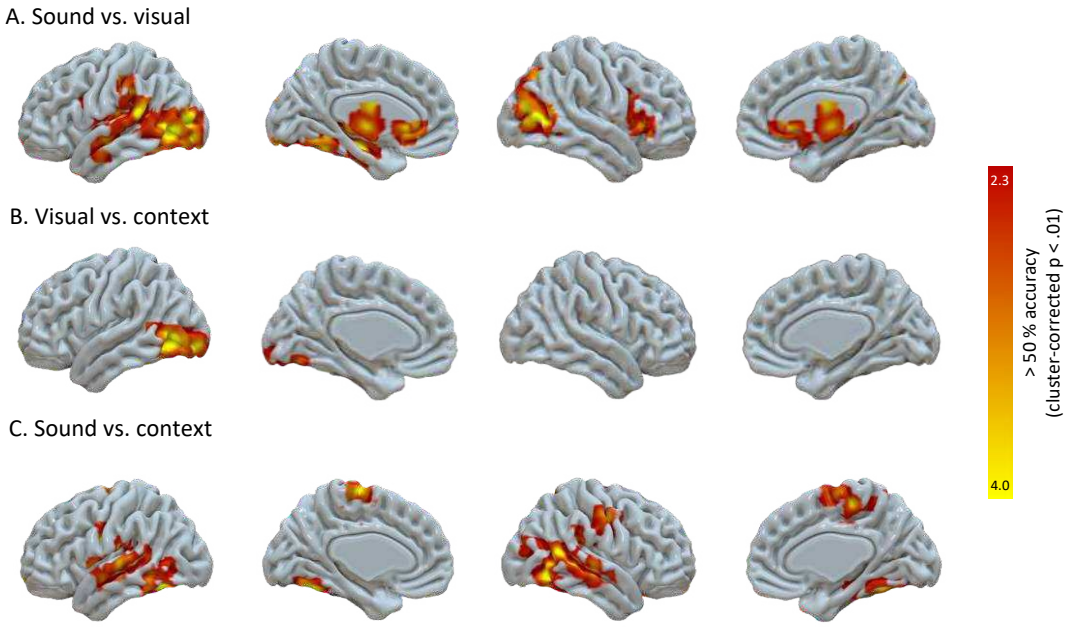

Figure 2. Results of the group-level whole-brain searchlight analysis with above-chance (50%) decoding projected in red (cluster-corrected $p < .01$). All panels reveal results from binary choice searchlight analyses decoding the content of thought while participants viewed visual and auditory noise. (A) Location of searchlights that could decode between thinking about the sound and thinking about the visual properties of concepts. (B) Location of searchlights that could decode between thinking about the visual properties of concepts and thinking about the same concepts in more complex contexts. (C) Location of searchlights that could decode between thinking about the sound of concepts and thinking about the same concepts in more complex contexts.

Next, we examined a visual vs. context classifier, which identified regions that could classify the difference between thinking about the visual properties of concepts and thinking about the same concepts in complex conceptual contexts. This whole-brain searchlight analysis revealed a large region in the left occipital lobe that could decode between visual and context conditions at above chance levels (50%, cluster-corrected $p < .01$) (Figure 2B; Table 2). Finally, we tested whether auditory vs. context conditions could be decoded. This whole-brain searchlight analysis revealed a set of clusters in bilateral auditory cortex extending along the superior temporal gyrus (STG) into ATL and posterior occipital-temporal cortex that could decode

between auditory and context conditions (50%, cluster-corrected p < .01) (Figure 2C; Table 2).

To identify regions that could consistently decode visual, auditory and context conditions, conjunction analyses were performed across the searchlight maps outlined in Figures 2A-C. The results of these conjunctions are presented in Figure 3A. For visual imagery, we looked at the conjunction of the two searchlight maps that involved decoding simple visual features (visual vs. auditory and visual vs. context). This revealed a left lateralized cluster in occipital pole extending into lateral occipital cortex, which reliably decoded the distinction between simple visual feature imagery and both of the other conditions. For auditory imagery, we looked at the conjunction of the two searchlight maps that involved decoding auditory properties (auditory vs. visual and auditory vs. context). This analysis revealed left hemisphere regions, including primary auditory cortex, STG, pMTG and occipital fusiform, that reliably decoded the distinction between simple auditory feature imagery and both of the other conditions. For imagery driven by complex conceptual contexts, we looked at the conjunction of the two searchlight maps that involved decoding context (visual vs. context and auditory vs. context), which produced a cluster in left lateral occipital cortex.

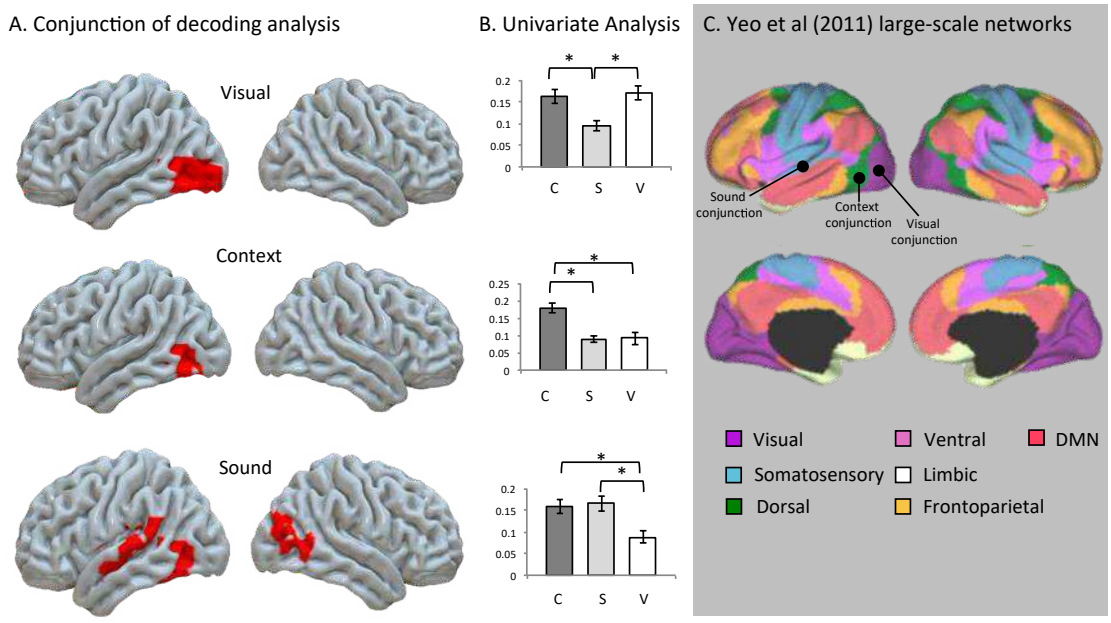

Figure 3. Panel A Represents brain regions where patterns of activity consistently informed the classifier for each of our three tasks (visual, context and sound). For visual patterns we

looked at the conjunction of the two searchlight maps that decoded visual properties (sound vs. visual *and* visual vs. context). For context patterns we looked at the conjunction of the two searchlight maps that decoded context properties (visual vs. context *and* sound vs. context). For sound patterns we looked at the conjunction of the two searchlight maps that decoded sound properties (sound vs. visual *and* sound vs. context). B. Shows the univariate percent signal change for each of our three conditions taken from a 6mm sphere centered on the peak conjunction point (visual [-48 -70 2], context [-48 -60 0], sound [-52 -8 -10]). (C = context, S = sound, V = visual). * Indicates a significant different between conditions ($p < .05$). The unthresholded maps for each condition have been uploaded to the Neurovault database and can be found here http://neurovault.org/collections/2671/. C. Grey panel illustrates the 7 core intrinsic networks identified by Yeo et al (2011); Dark purple = visual network, light blue = somatosensory network, dark green = dorsal network, light pink = ventral network, white = limbic network, yellow/orange = frontoparietal network (FPN) and red = default mode network (DMN). The black circles highlight where our peak conjunction sites fall with respect to these network. Our peak visual conjunction fell within the Visual network, peak context conjunction fell within the dorsal network and peak sound conjunction site within the somatosensory network.

The conjunction of the MVPA searchlight maps revealed regions of sensory cortex that could decode different types of imagery (Figure 3A). As an additional complementary analysis, the percentage signal change was extracted for each condition from each of the three conjunction sites by placing a 6mm sphere around the peak (Figure 3B). A 3 (conjunction site; visual, sound, conceptually-complex context) by 3 (imagery type: visual, sound, conceptually-complex context) repeated-measures ANOVA revealed no significant main effect of conjunction site ($F(2,36) = 0.48$, $p = .622$) or imagery type ($F(2,36) = 2.30$, $p = .114$; however there was a significant interaction between site and imagery type ($F(4,72) = 4.38$, $p = .003$). Planned comparisons in the form of repeated-measures t-tests revealed that our visual cluster showed significantly more activity for visual imagery than auditory imagery ($t(18) = 4.99$, $p < .001$) and for the context condition vs. auditory imagery ($t(18) = 4.61$, $p <.001$), but there was no significant difference between the visual and context conditions ($t(18) = .94$, $p = .36$). Our auditory cluster showed significantly more activity for auditory imagery than visual imagery ($t(18) = 4.64$, $p < .001$) and for the context condition vs. visual imagery ($t(18) = 5.602$, $p < .001$), but no significant difference between auditory and context imagery ($t(18) = -1.17$, $p = .25$). Finally, our context cluster revealed significantly more activity for the context condition compared to both visual ($t(18) = 5.56$, $p < .001$) and auditory imagery ($t(18) = 5.31$, p

< .001), but no significant difference between visual and auditory imagery conditions ($t(18) = -.03$, $p = .97$).

These univariate analyses demonstrate that regions that were able to classify particular aspects of internally-driven conceptual retrieval also showed a stronger BOLD response to these conditions – i.e., greater activation to visual or auditory imagery in 'visual' and 'auditory' classifier areas, and more activation to complex conceptual contexts in areas that could reliably classify this context condition. Regions that could decode visual and auditory imagery also responded to the context condition, consistent with the view that there is a multi-sensory response to complex conceptual contexts. Moreover, the context classifier region showed a response across both visual and auditory conditions, suggesting this region is transmodal; however, it also showed an increased response to imagery involving contexts, supporting the view that this region responds most strongly to the unique demands of the construction process imposed by this condition. Finally, to determine which distributed networks our conjunction findings fall within, we compared our results with seven large-scale networks as defined by Yeo et al (2011) (Figure 3C). Both visual and sound conjunction clusters fell predominantly within unimodal sensory networks (visual and somatosensory respectively), while our context conjunction site was located within the dorsal attentional network.

Given our priori predictions regarding heteromodal cortex (e.g., ATL), we interrogated candidate heteromodal regions within the auditory vs. visual classifier map. The brain regions labelled on Figure 4 are the peaks representing the highest decoding accuracy taken from Table 2, with the exclusion of peaks in unimodal cortex (determined by the conjunction results). This analysis included a distributed network of putative transmodal regions, including supramarginal gyrus extending into pMTG, ventrolateral ATL (aMTG and aITG), thalamus, anterior parahippocampal gyrus and anterior cingulate cortex (aCC) (Figure 4A). As before, the percent signal change was extracted from each of these regions by placing a 6mm sphere around each peak; SMG [-60 -42 16], aMTG [-56 -6 -18], aCC [-4 34-2], thalamus [-12 26 2] and aPG [-36 -18 -18]. A 5 (location; SMG, aMTG, aCC, thalamus, aPG) by 3 (imagery type: visual, sound, conceptually-complex context) repeated-measures ANOVA revealed no significant main effect of conjunction site ($F(4,72) = 0.34$, $p = .71$) or

imagery type (F(4,72) = 2.02, p = .131; nor was there a significant interaction between site and imagery type (F(8,144) = 2.65, p = .102). This equivalency across conditions is consistent with the characterization of these regions as transmodal. Finally, to quantify which intrinsic networks our clusters fall within we compared our results with seven large-scale networks as defined by Yeo et al (2011) (Figure 4B). The majority of clusters fell within transmodal cortices, including the default mode network and limbic system.

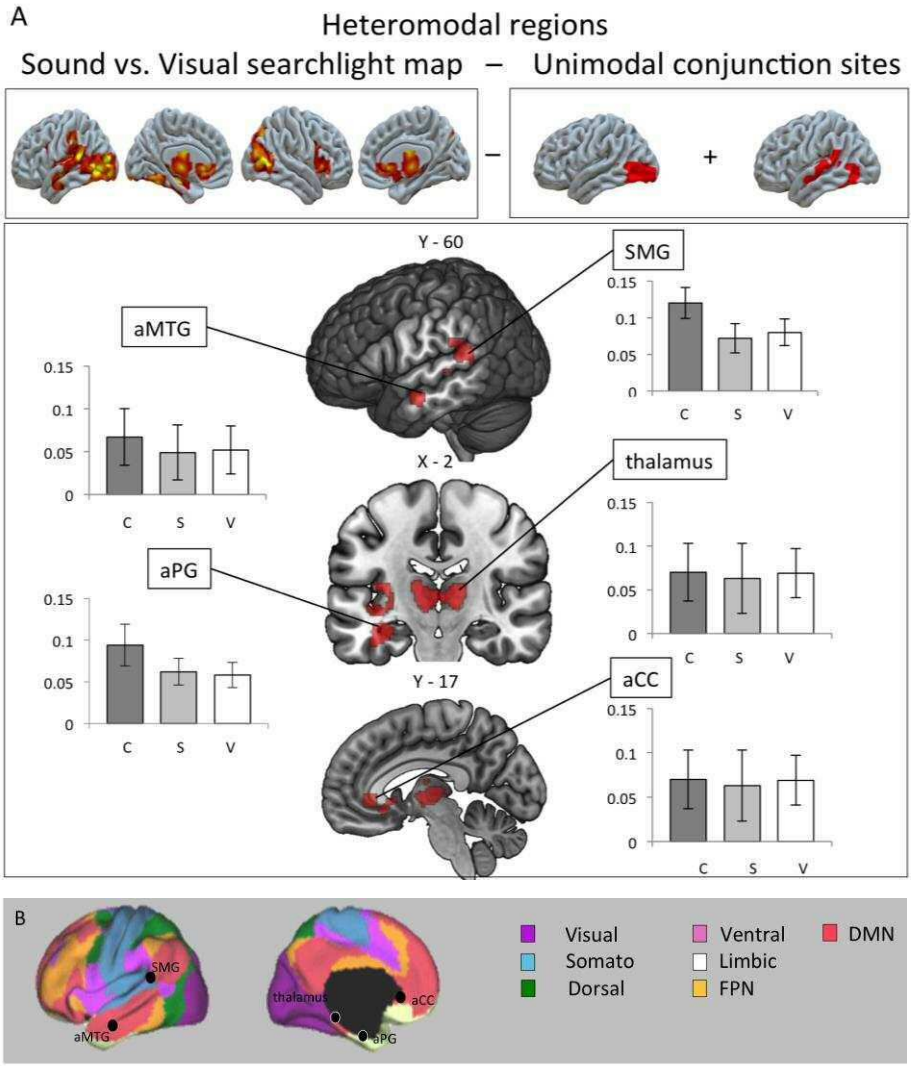

Figure 4. Heteromodal brain regions taken from the auditory vs. visual classifier map (Figure 2A). (A) Labelled regions highlight the peaks of decoding accuracy from Table 2 (excluding those peaks in unimodal cortex highlighted in our conjunction analysis for sound and visual imagination); SMG = supramarginal gyrus [-60 -42 16], aMTG = anterior middle temporal gyrus [-56 -6 -18], aCC = anterior cingulate cortex [-4 34-2], thalamus [-12 26 2], aPG= anterior parahippocampal gyrus [-36 -18 -18]. The bar graph shows the univariate percent signal change for each of our three conditions (C = context, S = sound, V = visual) extracted from a 6mm sphere centered on each labeled peak. There were no significant different

between conditions across any of our ROIs ($p > .05$). The unthresholded maps can be found here http://neurovault.org/collections/2671/. (B) Grey panel illustrates the 7 core intrinsic networks identified by Yeo et al (2011); Dark purple = visual network, light blue = somatosensory network, dark green = dorsal network, light pink = ventral network, white = limbic network, yellow/orange = frontoparietal network (FPN) and red = default mode network (DMN). The black circles highlight where our peak sites fall with respect to these network. SMG falls between ventral stream and somatomotor, aMTG, ACC fall within the default mode network, aPG falls within the limbic system. Subcortical regions (e.g., the thalamus) are not shown on the Yeo et al (2011) networks.

Intrinsic Connectivity

To provide a better understanding of the neural architecture that supported imagination in each condition, we explored the intrinsic connectivity of our unimodal conjunction sites (Figure 3) and transmodal sites (Figure 4) identified through MVPA, in resting-state fMRI. The results of the unimodal connectivity analysis are presented in Supplementary Table A2 (Figure 5A-C). For the visual and auditory conjunction sites, which peaked within visual and auditory cortex respectively, there was coupling beyond the sensory areas surrounding the seed regions, to include areas of transmodal cortex, including ATL, particularly the left medial surface, posterior middle temporal gyrus and precuneus. To quantify the interpretation of the functional connectivity of the visual, context and sound connectivity maps, we performed a decoding analysis using automated fMRI meta-analytic software NeuroSynth (right panel of Figure 5). Meta-analytic decoding of these spatial maps revealed domain specific networks and their associated function. The visual connectivity map correlated with terms related to visual processing (e.g., visual, objects), likewise our sound connectivity map correlated with terms related to auditory processing (e.g., speech, sound). The context connectivity map included both visual (e.g., objects) and higher-order terms (e.g., attention).

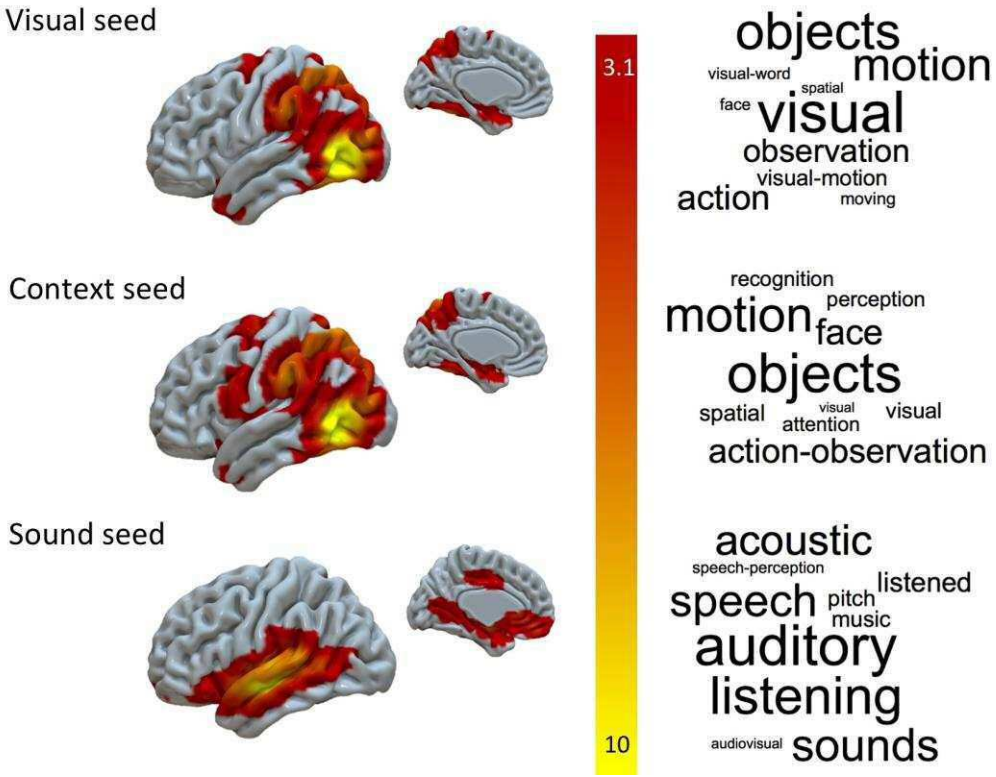

Figure 5. Resting state connectivity maps of unimodal regions projected on rendered brain, displaying left hemisphere and left medial view. Maps thresholded at z = 3.1, cluster corrected p < .01. Visual maps seeded from left inferior lateral occipital cortex [-48 70 -2]. Context maps seeded from left inferior lateral occipital cortex [-48 -60 0]. Sound maps seeded from left superior temporal gyrus [-52 -8 -10]. Word clouds represent the decoded function of each connectivity map using automated fMRI meta-analyses software (NeuroSynth, Yarkoni et al. 2011). This software computed the spatial correlation between each unthresholded zstat mask and every other meta-analytic map (n = 11406) for each term/concept stored in the database. The 10 meta-analytic maps exhibiting highest positive correlation for each sub-system was extracted, and the term corresponding to each of these meta-analyses is shown on the right. The font size reflects the size of the correlation. This allows us to quantify the most likely reverse inferences that would be drawn from these functional maps by the larger neuroimaging community.

Finally, the results of the heteromodal connectivity analysis are presented in Supplementary Table A2 (Figure 6A-B). Both our thalamus and SMG seed coupled

extensively with sensorimotor regions and core portions of the DMN (thalamus = angular gyrus and posterior cingulate cortex; SMG = middle temporal gyrus and ATL). The three other seeds (aMTG, anterior parahippocampal gyrus and anterior cingulate cortex) all coupled with core transmodal networks (DMN and limbic system). To aid the interpretation of these connectivity maps, we performed a decoding analysis using automated fMRI meta-analytic software NeuroSynth (right panel of Figure 6). The thalamus connectivity map correlated with terms related to task demands and multisensory properties (e.g., anticipation, motivation, somatosensory), likewise our SMG connectivity map correlated with terms related to sensory processing (e.g., speech, sound), while in contrast aMTG, aPG and aCC connectivity maps all correlated with terms related to memory retrieval (e.g., semantic, memory, encoding, DMN).

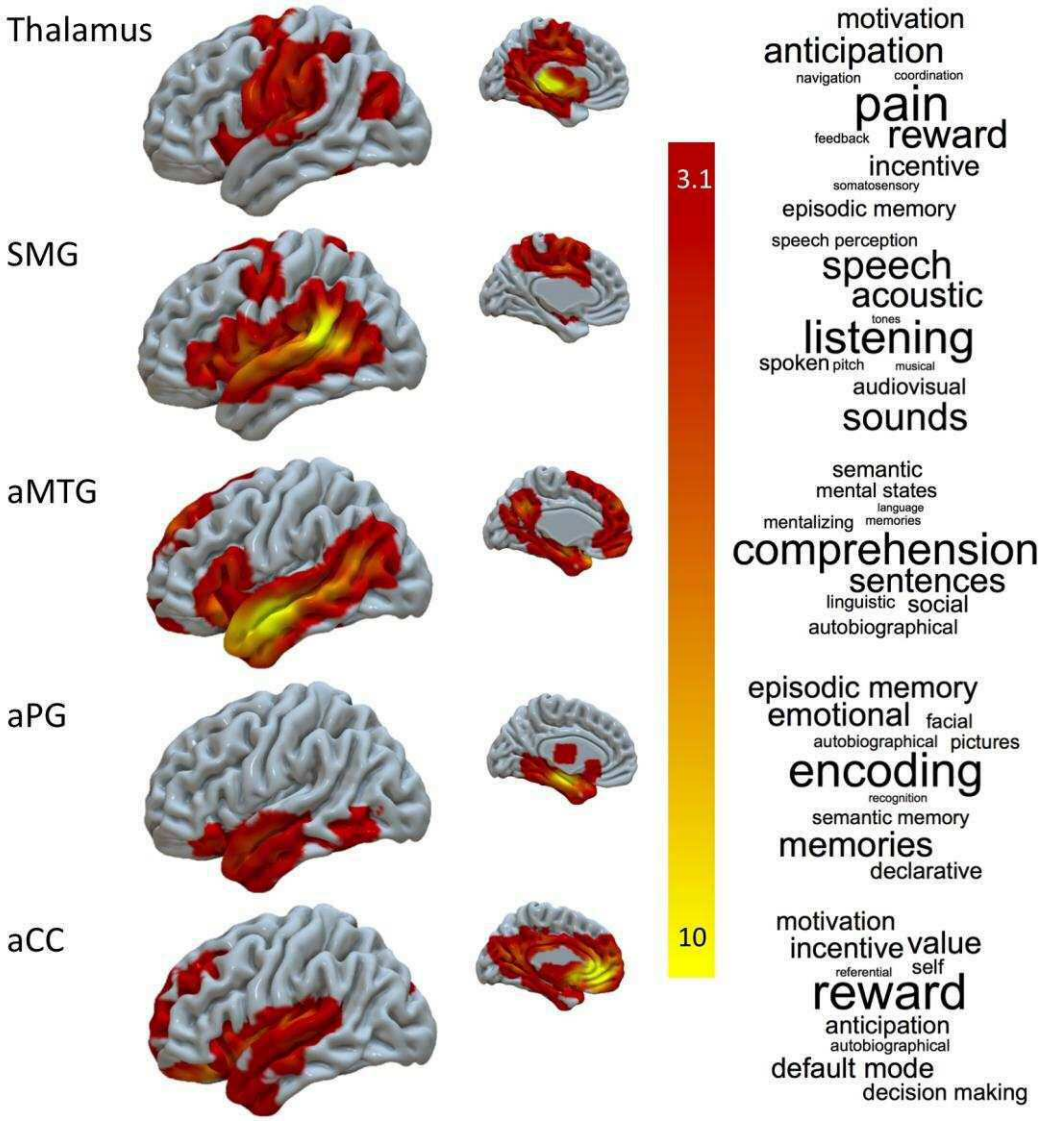

Figure 6. Resting state connectivity maps of heteromodal regions projected on rendered brain, displaying left hemisphere and left medial view. Maps thresholded at z = 3.1, cluster corrected p < .01. Thalamus maps seeded from [-48 -60 0]. Supramarginal gyrus (SMG) map seeded from [-48 -70 -2]. Anterior middle temporal gyrus (aMTG) seeded from [-56 -6 -18]. Anterior parahippocampal gyrus (aPG) seeded from [-36 -18 -18]. Anterior cingulate cortex (aCC) seeded from [-4 34-2]. Word clouds represent the decoded function of each connectivity map using automated fMRI meta-analyses software (NeuroSynth, Yarkoni et al. 2011). This software computed the spatial correlation between each unthresholded zstat mask and every other meta-analytic map (n = 11406) for each term/concept stored in the database. The 10 meta-analytic maps exhibiting highest positive correlation for each sub-system was extracted, and the term corresponding to each of these meta-analyses is shown on the right. The font size reflects the size of the correlation. This allows us to quantify the most likely reverse inferences that would be drawn from these functional maps by the larger neuroimaging community.

Discussion

Our study examined common and distinct components supporting conceptually-driven visual and auditory imagery. Multivariate whole-brain decoding identified aspects of secondary visual and auditory cortex (inferior lateral occipital cortex and superior temporal gyrus) in which the pattern of activation across voxels related to the modality of what was imagined. Using functional connectivity, we established that at rest these regions showed a pattern of differential connectivity with auditory or visual cortex, indicating that they reflected domain-specific aspects of imagination. We also identified several heteromodal regions (including ventrolateral ATL, anterior parahippocampal gyrus and anterior cingulate cortex) that were also able to decode the difference between thinking about what a concept looked like and what it sounded like. Finally, a region within the dorsal attention network (inferior lateral occipital cortex) was differentially recruited during imagination for more complex contexts and could reliably able to decode between all of our experimental conditions. Complementary investigation of the intrinsic connectivity of these regions confirmed their role in unimodal and heteromodal processing. These findings are consistent with the view that imagination emerges from a combined response within unimodal and transmodal regions.

The current fMRI study is one of only a few (e.g., Vetter et al., 2014) to identify patterns of activity in both visual and auditory association cortices that can reliably decode between different modalities of imagination (e.g., thinking about what a dog sounds like and what it looks like) within the same subjects. Our study is the first, to our knowledge, to investigate this issue whilst equating the visual and auditory input across our conditions. Typically neuroimaging studies of visual imagery have required participants to stare at a fixation cross while imagining an object, ensuring a consistent and simple visual input into the system (e.g., Albers et al. 2013; Dijkstra et al. 2017; Ishai et al. 2000; Lee, Kravitz & Baker, 2012; Reddy et al. 2010). In contrast, studies of auditory imagery typically require participants to imagine the sound of an object or piece of music in the presence of auditory input created by the scanner noise (e.g., Kraemer et al. 2005; Lima et al. 2015; 2016; Zattore & Halpern, 2005). In this study, we presented both visual and auditory random noise, providing more comparable visual and auditory baselines. This

methodological advance allows a purer test of common and distinct neural contributions to imagination within different modalities than has been possible in prior studies.

Domain specific contributions to imagination

Our study provided evidence that neural recruitment occurs in primary sensory regions in order to support modality-specific imagery. However, the highest decoding accuracy and the location of our imagination conjunctions fell within secondary sensory regions (superior temporal gyrus and inferior lateral occipital cortex respectively; Figure 3). Our functional connectivity analyses confirmed that although these regions fall outside of these systems as defined by Yeo and colleagues, at rest these regions are functionally coupled to primary visual and auditory cortex respectively. These findings are in line with prior decoding and fMRI studies that have highlighted the relationship between imagery and secondary sensory regions (Albers et al. 2013; de Borst & de Gelder, 2016; Coutanche & Thompson-Schill, 2014; Chen et al. 1998; Daselaar et al. 2010; Halpern et al. 2004; Ishai et al. 2000; Lee et al. 2012; Reddy et al. 2010; Stokes et al. 2009; Vetter et al. 2014; Zvyagintsev et al. 2013). Interestingly, our results are consistent with the 'anterior shift' noted by Thompson-Schill (2003). She found that areas activated by semantic processing are not isomorphic to those used in direct experience, but rather are shifted anterior to those areas (for a wider review see Chatterjee, 2010; Binder & Desai, 2011; McNorgan et al. 2011; Meteyard et al. 2012).

Our whole-brain searchlight analysis revealed patterns of activity supporting modality-specific imagination that extended beyond sensory cortex into semantic regions, including ATL (MTG, ventral and medial portions) and anterior cingulate cortex (see Figure 4). Functional connectivity analysis indicated that the majority of these regions showed extensive connectivity to other temporal lobe regions, encompassing both medial and lateral sites. Three of these regions also showed pre-frontal connectivity, primarily with connections to regions of the default mode network (anterior IFG and ventral and dorsal medial prefrontal cortex). Together this pattern of functional connectivity, suggests that these regions form a common

network in the temporal lobe, and at least some of these regions are closely allied at rest with regions within the default mode network.

Domain general contributions to imagination

We found a cluster in left inferior lateral occipital cortex (LOC) that showed stronger activation in the context condition. This region was able to classify the distinction between all three conditions. Left lateral occipital cortex is traditionally thought to support visual perception. However, this region predominantly falls within the dorsal attention network, as opposed to the visual network (Yeo et al. 2011). While this "task-positive" network usually responds to demanding, externally-presented decisions (for review see Corbetta & Shulman, 2002), in this study we see engagement in a task in which imagery is being generated internally from memory. This pattern of results demonstrates that imagery not only recruits transmodal regions associated with memory but also sites implicated in attention, when the features that are being retrieved have to be shaped to suit the context, and/or when complex patterns of retrieval are required. One caveat is that our current experimental paradigm does not allow us to establish if this response in LOC is driven by the need to generate rich heteromodal content (i.e., 'dog races' can envision the sound of a crowd cheering *and* the visual properties of a race track), or the requirement to steer retrieval away from dominant features to currently-relevant information (since the fact that dogs go for walks is not pertinent to 'dog races', and might need to be suppressed to allow contextually-relevant information to come to the fore). Nevertheless, the findings do suggest that this specific region plays a greater role in supporting imagery of complex multimodal contexts as opposed to single features.

Seeding from our "heteromodal" MVPA sites highlighted extensive functional coupling with core transmodal networks including DMN and limbic systems (see Figure 6; Margulies et al. 2016; Mesulam, 1989; Yeo et al. 2011). Meta-analytic decomposition of these maps returned terms related to memory retrieval (e.g., semantic, memory, encoding, DMN). In addition, two of these sites (thalamus and SMG) also coupled to somatosensory and attentional networks. Thalamic influence

has been previously reported during multisensory interplay (Driver & Noesselt, 2008) and its role in multimodal processing may explain why this region could decode between visual and auditory forms of imagination. Moreover, it has recently been suggested that SMG is crucial in the construction of mental representations (Benedek et al. 2017). As this region is connected to both attention and sensory networks, our findings converge with previous evidence suggesting that SMG integrates memory content in new ways and supports executively demanding mental simulations (Benedek et al. 2014; 2017; Fink et al. 2010).

6. Conclusion

In this investigation of semantic retrieval in the absence of meaningful stimuli in the external environment, we found extensive recruitment of sensory cortex, which was modulated by the modality of imagination required by the task. We also observed a role for transmodal brain regions in supporting internally-generated conceptual retrieval. These findings are consistent with the view that different types of imaginative thought depend upon patterns of common and distinct neural recruitment that reflect the respective contributions of modality specific and modality invariant neural representations.

References

Addis, D. R., Pan, L., Vu, M. A., Laiser, N., & Schacter, D. L. (2009). Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering. Neuropsychologia, 47(11), 2222-2238.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. Neuropsychologia, 45(7), 1363-1377.

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. Current Biology, 23(15), 1427-1431.

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: development, cognitive functions, phenomenology, and neurobiology. Psychological bulletin, 141(5), 931.

Amedi, A., Malach, R., & Pascual-Leone, A. (2005). Negative BOLD differentiates visual imagery and perception. Neuron, 48(5), 859-872.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. Psychological Review, 85(4), 249.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. Neuron, 65(4), 550-562.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. Annals of the New York Academy of Sciences, 1316(1), 29-52.

Antrobus, J. S., Singer, J. L., & Greenberg, S. (1966). Studies in the stream of consciousness: experimental enhancement and suppression of spontaneous cognitive processes. Perceptual and Motor Skills, 23, 399-417.

Bajada, C. J., Jackson, R. L., Haroon, H. A., Azadbakht, H., Parker, G. J., Lambon Ralph, M. A., & Cloutman, L. L. (2017). A graded tractographic parcellation of the temporal lobe. NeuroImage.

Baron, S. G., & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. Neuroimage, 55(4), 1847-1852.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. Behavioral and brain sciences, 22(4), 637-660.

Bemis, D. K., & Pylkkänen, L. (2012). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. Cerebral Cortex, 23(8), 1859-1873.

Barsalou, L. W. (2008). Grounded cognition. Annu. Rev. Psychol., 59, 617-645.

Benedek, M., Jauk, E., Fink, A., Koschutnig, K., Reishofer, G., Ebner, F., & Neubauer, A. C. (2014). To create or to recall? Neural mechanisms underlying the generation of creative new ideas. NeuroImage, 88, 125-133.

Benedek, M., Schues, T., Beaty, R. E., Jauk, E., Koschutnig, K., Fink, A., & Neubauer, A. (2017). To create or to recall original ideas: Brain processes associated with the imagination of novel object uses. *Cortex.*

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage, 37(1), 90-101.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. Trends in cognitive sciences, 15(11), 527-536.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex, 19(12), 2767-2796.

Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C., ... & Koenig, K. (2011). Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. Neuroimage, 54(2), 1465-1475.

Braga, R. M., Sharp, D. J., Leeson, C., Wise, R. J., & Leech, R. (2013). Echoes of the brain within default mode, association, and heteromodal cortices. Journal of Neuroscience, 33(35), 14031-14039.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. Trends in cognitive sciences, 11(2), 49-57.

Buckner, R. L., & Krienen, F. M. (2013). The evolution of distributed association networks in the human brain. Trends in Cognitive Sciences, 17(12), 648-665.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: mental imagery of complex sounds. Neuroimage, 26(4), 1119-1127.

Chatterjee, A. (2010). Disembodying cognition. Language and cognition, 2(1), 79-116.

Chen, W., Kato, T., Zhu, X. H., Ogawa, S., Tank, D. W., & Ugurbil, K. (1998). Human primary visual cortex and lateral geniculate nucleus activation during visual imagery. Neuroreport, 9(16), 3669-3674.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. Proceedings of the National Academy of Sciences, 106(21), 8719-8724.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nature reviews. Neuroscience, 3(3), 201.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Creating concepts from converging features in human cortex. Cerebral Cortex, 25(9), 2584-2593. Chicago

Daselaar, S. M., Porat, Y., Huijbers, W., & Pennartz, C. M. (2010). Modality-specific and modality-independent components of the human imagery system. Neuroimage, 52(2), 677-685.

Davey, J., Thompson, H. E., Hallam, G., Karapanagiotidis, T., Murphy, C., De Caso, I., ... & Jefferies, E. (2016). Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. NeuroImage, 137, 165-177.

de Borst, A. W., & de Gelder, B. (2016). fMRI-based multivariate pattern analyses reveal imagery modality and imagery content specific representations in primary somatosensory, motor and auditory cortices. Cerebral Cortex, 1- 15.

Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. J., & Friston, K. (2017). Distinct top-down and bottom-up brain connectivity during visual perception and imagery. Scientific Reports, 7.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron, 57(1), 11-23.

Farah, M. J. (1984). The neurological basis of mental imagery: A componential analysis. Cognition, 18(1), 245-272.

Fink, A., Grabner, R. H., Gebauer, D., Reishofer, G., Koschutnig, K., & Ebner, F. (2010). Enhancing creativity by means of cognitive stimulation: Evidence from an fMRI study. NeuroImage, 52(4), 1687-1695.

Friedman, L., Glover, G. H., & Fbirn Consortium. (2006). Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal- to-fluctuation-noise-ratio (SFNR) differences. Neuroimage, 33(2), 471-481.

Gabrieli, J. D., Brewer, J. B., Desmond, J. E., & Glover, G. H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. Science, 276(5310), 264-266.

Halpern, A. R. (2001). Cerebral substrates of musical imagery. Annals of the New York Academy of Sciences, 930(1), 179-192.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. Cerebral cortex, 9(7), 697-704.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia, 42(9), 1281-1292.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics, 7(1), 37-53.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. Trends in cognitive sciences, 11(7), 299-306.

Hauk, O., & Tschentscher, N. (2013). The body of evidence: what can neuroscience tell us about embodied semantics?. Frontiers in Psychology, 4.

Hertz, U., & Amedi, A. (2010). Disentangling unisensory and multisensory components in audiovisual integration using a novel multifrequency fMRI spectral analysis. Neuroimage, 52(2), 617-632.

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. Neuron, 28(3), 979-990.

Jackson, R. L., Hoffman, P., Pobric, G., & Lambon Ralph, M. A. (2016). The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions. Journal of Neuroscience, 36(5), 1490-1501.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage, 17(2), 825-841.

Jezzard, P., & Clare, S. (1999). Sources of distortion in functional MRI data. Human brain mapping, 8(2-3), 80-85.

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. Psychological science, 18(7), 614-621.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. Cortex, 48(7), 805-825.

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. Science, 330(6006), 932-932.

Knauff, M., Kassubek, J., Mulack, T., & Greenlee, M. W. (2000). Cortical activation evoked by visual mental imagery as measured by fMRI. Neuroreport, 11(18), 3957-3962.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. Nature reviews. Neuroscience, 2(9), 635.

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Ganis, G., ... & Alpert, N. M. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. Science, 284(5411), 167-170.

Kraemer, D. J., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: sound of silence activates auditory cortex. Nature, 434(7030), 158-158.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. Proceedings of the National academy of Sciences of the United States of America, 103(10), 3863-3868.

Lambon Ralph, M., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. Nat. Rev. Neurosci., 18, 42-55.

Lee, S. H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. Neuroimage, 59(4), 4064-4073.

Leech, R., Braga, R., & Sharp, D. J. (2012). Echoes of the brain within the posterior cingulate cortex. Journal of Neuroscience, 32(1), 215-222.

Lewis-Peacock, J. A., & Norman, K. A. (2013). Multi-voxel pattern analysis of fMRI data. The cognitive neurosciences, 911-920.

Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., ... & Warren, J. E. (2015). Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems. Cerebral cortex, 25(11), 4638-4650.

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., ... & Jefferies, E. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. Proceedings of the National Academy of Sciences, 113(44), 12574-12579.

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. Science, 315(5810), 393-395.

McNorgan, C., Reid, J., & McRae, K. (2011). Integrating conceptual knowledge within and across representational modalities. Cognition, 118(2), 211-233.

Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: neuronal interactions during perception and imagery. Cerebral cortex, 14(11), 1256-1265.

Mesulam, M. (2012). The evolving landscape of human cortical connectivity: facts and inferences. Neuroimage, 62(4), 2182-2189.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. Cortex, 48(7), 788-804.

Miller, B. T., & D'Esposito, M. (2005). Searching for "the top" in top-down control. Neuron, 48(4), 535-538.

Murphy, C., Jefferies, E., Rueschemeyer, S. A., Sormaz, M., Wang, H. T., Margulies, D. S., & Smallwood, J. (2018). Distant from input: Evidence of regions within the default mode network supporting perceptually-decoupled and conceptually-guided cognition. *NeuroImage*.

Murphy, C., Rueschemeyer, S. A., Watson, D., Karapanagiotidis, T., Smallwood, J., & Jefferies, E. (2017). Fractionating the anterior temporal lobe: MVPA reveals differential responses to input and conceptual modality. NeuroImage, 147, 19-31.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in cognitive sciences, 10(9), 424-430.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. Nature reviews. Neuroscience, 8(12), 976.

Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. Cognitive Neuropsychology, 19(7), 603-639.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. Trends in cognitive sciences, 17(9), 458-470.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. Proceedings of the National Academy of Sciences, 98(2), 676-682.

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: decoding category information during mental imagery. NeuroImage, 50(2), 818-825.

Reilly, J., Garcia, A., & Binney, R. J. (2016). Does the sound of a barking dog activate its corresponding visual form? An fMRI investigation of modality-specific semantic access. Brain and language, 159, 45-59.

Rice, G. E., Hoffman, P., Lambon Ralph, M. A., & Matthew, A. (2015). Graded specialization within and between the anterior temporal lobes. Annuals of the New York Academy of Sciences, 1359(1), 84-97.

Rueschemeyer, S.-A., Eckmann, M., van Ackeren, & Kilner, J. (2014). Observing, performing and understanding actions: Revisiting the role of cortical motor areas in processing of action words. Journal of Cognitive Neuroscience, 26(8), 1644-1653.

Rugg, M. D., & Vilberg, K. L. (2013). Brain networks underlying episodic memory retrieval. Current opinion in neurobiology, 23(2), 255-260.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philosophical Transactions of the Royal Society B: Biological Sciences, 362(1481), 773-786.

Shulman, G. L., Corbetta, M., Buckner, R. L., Raichle, M. E., Fiez, J. A., Miezin, F. M., & Petersen, S. E. (1997). Top-down modulation of early sensory cortex. Cerebral cortex (New York, NY: 1991), 7(3), 193-206.

Singer, J. L. (1966). Daydreaming: An introduction to the experimental study of inner experience.

Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. Cerebral cortex, 15(10), 1570-1583.

Smallwood, J., Karapanagiotidis, T., Ruby, F., Medea, B., de Caso, I., Konishi, M., ... & Jefferies, E. (2016). Representing representation: Integration between the

temporal lobe and the posterior cingulate influences the content and form of spontaneous thought. PloS one, 11(4), e0152272.

Smith, S. M. (2002). Fast robust automated brain extraction. Human brain mapping, 17(3), 143-155.

Sormaz, M., Jefferies, E., Bernhardt, B. C., Karapanagiotidis, T., Mollo, G., Bernasconi, N., ... & Smallwood, J. (2017). Knowing what from where: Hippocampal connectivity with temporoparietal cortex at rest is linked to individual differences in semantic and topographic memory. NeuroImage, 152, 400-410.

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape- specific population codes in visual cortex during mental imagery. Journal of Neuroscience, 29(5), 1565-1572.

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. Proceedings of the National Academy of Sciences, 104(2), 642-647.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring “how” from “where”. Neuropsychologia, 41(3), 280-292.

Van Ackeren, M. & Rueschemeyer, S.-A. (2014). Theta power and beta coherence predict multimodal semantic integration at different cortical scales. PLOS ONE, 9(7), e101042.

Vatansever, D., Bzdok, D., Wang, H. T., Mollo, G., Sormaz, M., Murphy, C., ... & Jefferies, E. (2017). Varieties of semantic cognition revealed through simultaneous decomposition of intrinsic brain connectivity and behaviour. *Neuroimage*, *158*, 1-11.

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. Current Biology, 24(11), 1256-1262.

Visser, M., Jefferies, E., & Lambon Ralph, M. (2010). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. Journal of cognitive neuroscience, 22(6), 1083-1094.

Wang, H. T., Poerio, G., Murphy, C., Bzdok, D., Jefferies, E., & Smallwood, J. (2017). Dimensions of experience: exploring the heterogeneity of the wandering mind. *Psychological science*, 0956797617728727.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. Nature methods, 8(8), 665-670.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... & Fischl, B. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of neurophysiology, 106(3), 1125-1165.

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: musical imagery and auditory cortex. Neuron, 47(1), 9-12.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE transactions on medical imaging, 20(1), 45- 57.

Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., & Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. European Journal of Neuroscience, 37(9), 1421-1434.

Supplementary Material

Supplementary material A1: Description of pilot experiment.

For the visual and context trials (car visual, car context, dog visual, dog context), a pictorial target was used (e.g., a picture of a car tyre for the car visual condition). For auditory trials, a sound target was used (e.g., a dog barking). On each trial of this behavioral pilot, participants were presented with both visual and auditory noise. One of two target types were then superimposed over the visual and auditory noise: (i) image targets and (ii) sound targets. For image targets, 150 different images were presented centrally to participants. There were 30 images for each of the following experimental conditions: Dog Visual-Features (e.g., dog paw), Cars Visual-Features (e.g., car tyre), Dog Contexts (e.g., race dosg) and Car Contexts (e.g., race car) and an additional 30 catch-trials (that did not represent any of the experimental conditions). Each item emerged through the noise by adjusting the opacity of the image from 0 (transparent) to 1 (opaque) in increments of 0.025 every 150ms. For sound trials, 90 different sounds were presented binaurally to participants. There were 30 sounds for each of following sound experimental conditions: Dog Sounds (e.g., barking), Car Sounds (e.g., breaks screeching) and an additional 30 catch-trials (that did not represent the other experimental conditions). All sound trials were modified to have the same average amplitude. Each sound emerged through noise by adjusting the volume from 0 to 1 in increments of 0.10. Each sound was played in full before the volume increased (the maximum length of any of the sound trials was 600ms).

For this pilot test, participants were instructed to respond with a button-press when they could identify the image or sound emerging through the noise. Images were presented first (for all image-based conditions), followed by sound trials. The order of presentation of individual image and sound trials was randomized across participants. To ensure that participants were accurately identifying the images and sounds, on 10% of trials participants were also required to type what they had seen or heard. The average detection time across all participants was calculated for every image and sound trial. Ten images were then selected for each of our six experimental conditions (Dog Visual-Features, Car Visual-Features, Dog Sound, Car Sound, Dog Context and Car Context) based on statistically similar

reaction times (RTs) for detecting the item emerging through noise. Images were detected on average at 2861ms and sounds at 2912ms (see Table 1). These timings were used in the fMRI experiment to ensure that the in-scan detection task would be challenging enough to engage all participants. The fMRI scan therefore allowed 3000ms for participants to detect an item emerging through noise.

Table A2. List of stimuli

	Sound	Visual	Context
Dog	“Sound Dog”	Visual Dog”	“Race Dog”
			“Dangerous Dog”
			“Old Dog”
			“New Dog”
			“Muddy Dog”
			“Clean Dog”
			“Abandoned Dog”
			“Family Dog”
Car	“Sound Car”	Visual Car”	“Race Car”
			“Dangerous Car”
			“Old Car”
			“New Car”
			“Muddy Car”
			“Clean Car”
			“Abandoned Car”
			“Family Car”

Footnote: Prompt for each experimental conditions depicted in “ “.

Table A3. Coordinates of peak clusters in the resting-state connectivity analyses.

Seed Region	Cluster	Cluster Extent	Z-score	x	Y	z
Context seed	*Increased Correlation*					
	L. Lateral occipital cortex, inferior division	15566	16.4	-50	-64	0
	L. Superior frontal gyrus	566	8.18	-22	-8	54
	R. Planum polare	256	5.45	42	-10	-8
	Reduced Correlation					
	R. Lingual gyrus	6653	7.27	4	-88	14
	R. Anterior cingulate gyrus	5584	7.14	6	26	30
	R. Insular Cortex	2324	6.46	38	14	-10
	L. Postcentral Gyrus	340	4.75	-60	-6	14
	L. Frontal Pole	296	4.43	-36	50	12
	R. Lateral occipital pole, superior division	265	4.8	48	-64	48
Visual seed	*Increased Correlation*					
	L. Lateral occipital cortex, inferior division	7797	15.3	-48	-68	0
	R. Lateral occipital cortex, inferior division	6793	10.9	50	-64	2
	L. Hippocampus	346	5.29	-20	-10	-20
	L. Superior Frontal gyrus	342	7.47	-22	-8	54
	Reduced Correlation					
	R. Lingual gyrus	6688	7.35	4	-70	-4
	R. Insular cortex	2463	6.31	40	12	-6
	R. Paracingulate gyrus	2369	6.85	10	22	34
	R. Frontal pole	2270	6.17	38	40	18
	L. Insular cortex	856	5.42	-36	4	2
	L. Frontal pole	388	5.25	-34	50	8
	R. Posterior cingulate gyrus	354	4,59	2	-32	26
Sound seed	*Increased Correlation*					
	L. Superior temporal gyrus	17702	15.8	-46	-10	-6

	R. Intracalcarine cortex	614	5.45	20	-62	10
	L. Lingual gyrus	564	5.27	-16	-51	0
	R. Anterior cingulate gyrus	511	4.54	6	-14	42
	Reduced Correlation					
	R. Thalamus	1961	6.28	16	-14	10
	L. Superior frontal gyrus	1685	4.98	-20	10	62
	R. Cerebellum	1445	5.58	36	-58	-48
	L. Cerebellum	1187	5.6	-36	-50	-48
	L. Lateral occipital cortex, superior division	1110	5.16	-26	-72	30
	R. Lateral occipital cortex, superior division	670	5.86	26	-78	34
	R. Superior frontal gyrus	571	5.68	26	-4	52
	L. Lateral occipital cortex, inferior division	364	4.38	-48	-78	-12
	L. Frontal pole	291	5.07	-26	54	2
Thalamus	*Increased Correlation*					
	L. Thalamus	22269	17.1	-12	-26	2
	L. Lateral occipital cortex, superior division	270	5.17	-42	-72	24
	Reduced Correlation					
	L. Cerebellum	19581	7.66	-40	-74	-32
	L. Frontal pole	786	5.58	-26	54	20
	L. Planum polare	258	6.29	-44	-10	-12
SMG	*Increased Correlation*					
	L. Supramarginal gyrus, posterior division	9745	15.1	-60	-42	16
	R. Planum temporale	7485	8.64	52	-32	18
	L. Cingulate gyrus, anterior division	4128	7.26	-6	-12	36
	L. Precentral gyrus	330	4.96	-46	-8	44
	R. Cerebellum	289	5.43	26	-72	-56
	Reduced Correlation					
	R. Lateral occipital cortex, superior division	6353	7.45	26	-66	52
	L. Lateral occipital cortex, superior division	3955	6.63	-28	-60	46

	R. Middle frontal gyrus	1529	6.19	38	8	60
	L. Superior frontal gyrus	552	5.95	-26	18	58
	L. Cerebellum	245	5.23	-44	-68	-46
aMTG	*Increased Correlation*					
	L. Middle temporal gyrus, anterior division	10430	15.3	-56	-6	-18
	R. Middle temporal gyrus, posterior division	7048	10.2	50	-12	-16
	L. Posterior cingulate gyrus	2696	7.34	-8	-54	32
	L. Superior frontal gyrus	1606	7.69	-8	52	32
	L. Frontal pole	821	5.94	-6	56	-14
	Reduced Correlation					
	R. Frontal pole	3034	6.85	46	46	12
	L. Frontal pole	1397	6.56	-46	42	16
	R. Angular gyrus	1178	6.25	42	-52	50
	L. Supramarginal gyrus, posterior division	1158	6.90	-50	-42	44
	L. Cerebellum	1108	6.15	-32	-70	-34
	R. Paracingulate gyrus	781	6.91	4	20	42
	R. Superior frontal gyrus	734	5.47	20	16	56
	R. Cerebellum	648	5.69	40	-56	-54
	L. Superior frontal gyrus	490	5.38	-24	2	56
	L. Lingual gyrus	337	4.54	-2	-82	-24
	Thalamus	246	4.64	0	-4	2
	L. Precuneous	210	4.87	-14	-74	42
aPG	*Increased Correlation*					
	L. Parahippocampal gyrus, anterior division/temporal fusiform cortex	15370	15.6	-36	-16	-18
	L. Thalamus	207	4.88	-2	-14	6
	Reduced Correlation					
	R. Middle frontal gyrus	7768	7.09	34	16	50
	R. Lateral occipital cortex, superior division	2232	6.83	46	-62	30
	Intracalcarine cortex	2115	4.71	12	-82	4

	L. Middle frontal gyrus	1893	5.90	-34	2	50
	L. Angular gyrus	1016	5.50	-54	-58	36
	L. Thalamus	659	5.79	-8	-14	-2
ACC	*Increased Correlation*					
	L. Cingulate gyrus, anterior division	28384	15.4	-4	34	-2
	R. Lateral occipital cortex, superior division	315	5.56	52	-68	20
	L. Middle frontal gyrus	272	6.21	-24	32	34
	Reduced Correlation					
	R. Cerebellum	7277	7.76	12	-80	-34
	L. Inferior frontal gyrus, pars opercularis	3364	6.43	-54	14	20
	R. Inferior frontal gyrus, pars opercularis	2065	5.47	52	16	18
	L. Lateral occipital cortex, superior division	1782	6.82	-30	-64	40
	R. Lateral occipital cortex, superior division	750	4.93	36	-66	46
	L. Paracingulate gyrus	468	4.61	-4	28	44

Footnote: The table shows peak clusters in the resting-state connectivity analysis from eight seed regions. Three "unimodal" regions; context seed [-48 60 0], visual seed [-48 -70 -2] and sound seed [52 -8 -10]. Results are thresholded at p<.01 (cluster corrected). Five "heteromodal" regions; Thalamus seed [-48 -60 0], supramarginal gyrus (SMG) seed [-48 -70 -2], anterior middle temporal gyrus (aMTG) seed [-56 -6 -18], anterior parahippocampal gyrus (aPG) seed [-36 -18 -18] and anterior cingulate cortex (aCC) seed [-4 34-2]. L=left, R=right.